Comparison of luminance based metrics in different lighting conditions

Wienold, J.; Kuhn, T.E.; Christoffersen, J.; Sarey Khanie, Mandana; Andersen, M.

Publication date:
2017

Document Version
Peer reviewed version

COMPARISON OF LUMINANCE BASED METRICS IN DIFFERENT LIGHTING CONDITIONS

Wienold,J.[1], Kuhn, T.E.[2], Christoffersen, J. [3] ,Sarey Khanie, M.[1,4], Andersen, M.[1]
[1]École Polytechnique Fédérale de Lausanne (EPFL), [2]Fraunhofer Institute for Solar Energy Systems ISE, Freiburg, [3]VELUX Group, Denmark, [4]Technical University of Denmark DTU
jan.wienold@epfl.ch

Abstract

In this study, we evaluate established and newly developed metrics for predicting glare using data from three different research studies. The evaluation covers two different targets: 1. How well the user's perception of glare magnitude correlates to the prediction of the glare metrics? 2. How well do the glare metrics describe the subjects' disturbance by glare? We applied Spearman correlations, logistic regressions and an accuracy evaluation, based on an ROC-analysis. The results show that five of the twelve investigated metrics are failing at least one of the statistical tests. The other seven metrics CGI, modified DGI, DGP, Ev, average Luminance of the image L_{avg}, UGP and UGR are passing all statistical tests. DGP, CGI, DGI_mod and UGP have largest AUC and might be slightly more robust. The accuracy of the predictions of afore mentioned seven metrics for the disturbance by glare lies in the range of 75-83% and does not confirm findings from other studies stating a poor performance of existing glare metrics.

Keywords: Daylight, glare, glare perception, user assessments

1 Introduction and Objectives

In recent years user assessment studies on daylight-induced visual discomfort for application mainly in office buildings have been conducted more extensively [1-5]. These studies are proposing different metrics to predict discomfort caused by the luminance distribution in the field of view. However, in several cases only minimum changes have been considered in the experimental set up as far as the luminous environment is concerned, thereby limiting the applicability of the metrics for broader ranges of luminance distribution. The question remains how the proposed metrics will perform when the lighting conditions are different from the ones in which they were developed.

For instance, K. Van Den Wymelenberg et al. [5,6] concluded, that the overall performance of existing glare metrics is rather poor. They based their conclusions on a Pearson correlation between ordinal subjective response and the metric values. Hirning et al [7] found, that existing glare metrics generally underestimate the subjective glare response of users in the context of open-plan offices in green-buildings (located in Australia). Both studies suggest new metrics that attempt to improve their ability to predict glare sensation from subjects. However, an underlying question always remains: how do existing, revised and other newly published metrics perform in conditions in which they have not been developed?

In general, the variability between the individual subjects' perception of visual discomfort has been addressed by several authors [2,8,9] and might explain the low levels of correlations found. However, there are attempts not trying to explain individual differences in glare perception but describing the probability that a subject is being disturbed by glare [2]. This approach will be used as one of several evaluation methods in this study.

The objective of this study is to evaluate established and newly proposed glare prediction metrics by using data-sets that were not used to develop the metrics themselves. Several statistical analysis methods are here applied to evaluate glare prediction metrics by comparing them with the users' glare ratings emanating from three different research studies.

2 Investigated glare metrics

Among the developed metrics for predicting daylight-induced discomfort glare risk, Daylight Glare Index (**DGI**)[1] can be identified as an early attempt to express the magnitude of glare

perceived by humans. Although it was developed under artificial light it describes the glare magnitude caused of daylight from a window.

Daylight Glare Probability (**DGP**) [2], was the first metric developed under real daylight conditions where user assessment data were acquired in office like test rooms. Fisekis and Davis [4] suggested a modification of the DGI (named here in the following $\mathbf{DGI_{mod}}$). K. Van Den Wymelenberg et al. [5,6] suggest to use the average luminance in the 40°band ($\mathbf{L_{40°band\_avg}}$) and the standard deviation of window luminance ($\mathbf{L_{std\_window}}$) for the evaluation of visual discomfort. As an outcome of field-studies in green buildings in Australia, Hirning et. al. propose Unified Glare Probability (**UGP**) [7] for glare analysis in this kind of setting. Tokura et. Al. developed the predicted glare sensation vote (pgsv) [17] to describe glare from windows. In this study, we investigate a modified version of the pgsv for describing the saturation effect, published by Iwata et. al. [18] ($\mathbf{pgsv_{sat}}$).

Besides the above-mentioned metrics, photometric quantities such as average luminance in the field of view (FOV) $\mathbf{L_{avg}}$, the average Window luminance $\mathbf{L_{avg\_window}}$ and the vertical illuminance at eye level $\mathbf{E_v}$ will be used for the evaluations.

Also, two metrics for electric light glare evaluations will be applied in this study: CIE Glare Index (**CGI)** [14,15] and Unified Glare rating (**UGR)** [16].

The twelve above-mentioned metrics in **bold** are investigated in this study and are derived from calibrated high dynamic (HDR) fisheye images by evalglare [19] (version 1.31). The glare source detection mode was set to the recommended [2] setting to 5-times the task-area luminance.

3 Methodology

User assessment data and related high dynamic range (HDR) fisheye images from three different experiments conducted between 2003 and 2013 are used for this study. For each case, twelve glare metrics are calculated from the HDR images. For each case, the glare response of the subjects on an ordinal 4-point Likert-scale is available. The scale is the same for each study: imperceptible-noticeable-disturbing-intolerable.

The datasets are split up into different sub-datasets in order to guarantee a reliable statistical analysis (see chapter 3.2).

3.1 Study description

Experimental data from three different studies are used for this paper.

The experiments for the study “Ecco-Build” were conducted between 2003 and 2005 in Copenhagen (DK) and Freiburg (D). The experiments were conducted in an office-like setup. Three different window sizes and three different shading devices (white Venetian blinds, specular reflective blinds, foil shading) were used to create different lighting conditions. The experimental setup is described in detail in [2]. 348 of the 366 cases were used to develop the DGP metric. For comparing the DGP with the other metrics, the development-data are either excluded, or the results for the DGP evaluation are displayed in a separate colour.

The experiments for the study “Quanta” were conducted between 2008 and 2011 in Freiburg (D) in the same facility as in “Ecco-Build” and therefore the experimental setup is very similar. The main differences are that other shading devices were examined (fabric shading devices), the window size was fixed and that both rooms were used for the user assessments and the luminance cameras were placed besides the subject. Details are described in [10].

The “Gaze” study was conducted in 2012 in the same facility in Freiburg (D) and used no shading devices and had situations with and without the sun in the field of view. Details are described in [3]. An overview over the three studies is shown in table 1 including the number of subjects and cases for each study.

Table 1 – Overview of the three studies and their experimental setup.

Name	Location	Main publ.	Tested façade systems	No. Subjects	No. cases
Ecco-Build	Freiburg, Germany Copenhagen, Denmark	[2]	Three window sizes, three shading devices, two viewing directions	59 24	235 131
Quanta	Freiburg, Germany	[10]	Fabrics with and without view, venetian blinds	49	188
Gaze	Freiburg, Germany	[3]	Window not shaded	100	100
Total				**232**	**654**

The entire data-set contains therefore data from five different shading devices plus one widow configuration without shading. For a subset of the data (Ecco-Build) also the window size was varied for three shading devices. Therefore, the dataset has a large variety of lighting conditions (see figure 1).

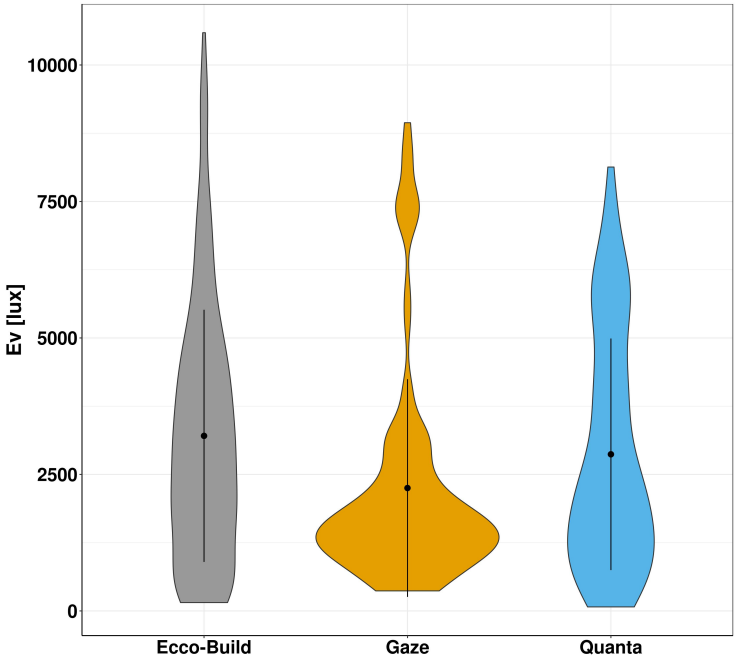


Figure 1 – Violin plot for the distribution of the vertical Illuminance at Eye level E_v. The mean value (dot) as well as the standard deviation (vertical line) are displayed as well.

3.2 Data selection

The data preparation and selection is adapted to the statistical evaluations. Several constraints are considered:

1. Some of the evaluations derive cut-off-values to be applied to the data later. The cut-off-point of a metric is defined as the value of the borderline of the metric, dividing “not-disturbed” from “disturbed”. This process is seen as modelling and therefore the data are split up into a dataset for the modelling (in the following called “training-dataset”, using 220 cases) and a dataset for the evaluation/comparisons (in the following called “testing-dataset”, using 195 cases). The split of the data was done randomly. This data selection procedure applies to the accuracy analysis and to the analysis of the probability of persons disturbed by glare.

2. A large part (348 of the 366 cases) of the Ecco-Build dataset was originally used to develop the DGP metric. Therefore, any comparison between the metrics considering the DGP is excluding this development-data in order not to bias the results. A combined-dataset of the three studies excluding this development data is called “non-dev dataset”. For the calculation of the cut-off-point (“training-dataset”), the development data are included.

3. For the Spearman correlations on the study data, all data are included in order to see the influence of the individual study on the correlation of the metrics. The correlation value of the DGP in that case cannot be compared to the other values and is therefore marked in grey.

3.3 Statistical methods

So far, there exists no single, commonly accepted statistical method to evaluate the performance of glare metrics. Already the word “performance” can be interpreted in different ways. What are the criteria to decide, if a metric “performs” well? For this study, we concentrate on two, in our opinion most important, questions:

1. How well does the metric describe the entire glare scale? This evaluation seems to be the most natural way looking at the data.

2. How well does the metric describe the probability, that a person is disturbed by glare? This kind of approach is a probabilistic one, which has been in use for many years in medical diagnostics and which was introduced to the glare evaluation field by [2].

To answer the first question, K. Van Den Wymelenberg et al. [5,6] applied Pearson’s correlation between ordinal subjective responses and the metric values. This statistical method delivers reliable results only when the distance between all the ordinal categories are known as to be the same. Typically, this is not the case and especially not for the underlying Likert-4-point scale of the subjective responses to glare of this study. For the evaluation of ordinal data the use of Spearman correlation is appropriate [12]. We apply this kind of evaluation within this study. Furthermore, the p-values of the Spearman statistics is evaluated.

To answer the second question, Rodrigez et. Al. [11] is using an epidemiological approach applying diagnostic accuracy methods (Sensitivity, Specificity, Positive Likelihood Ratio, Negative Likelihood Ratio, Youden's Index, ROC Square Distance). In [2], the p-values of a logistic regression as well as a squared Pearson correlation on the probability data is applied. For this study, we decided to apply parts of the diagnostic accuracy method as well as the logistic regression and the squared Pearson correlation on the probability.

The evaluations are explained in 3.31 – 3.3.6:

3.3.1 Spearman-correlation

The Spearman rank correlation ρ is a non-parametric test to measure the strength of the relationship between paired data. Differently to the Pearson correlation, the underlying independent variables don’t need to be of numerical or equidistant-ordinal nature [12]. The higher the value the stronger the correlation between the variables.

For this study, we derive ρ and the related p-value. The p-value is also compared to Bonferroni-corrected significance values.

3.3.2 Bonferroni correction for the significance-levels

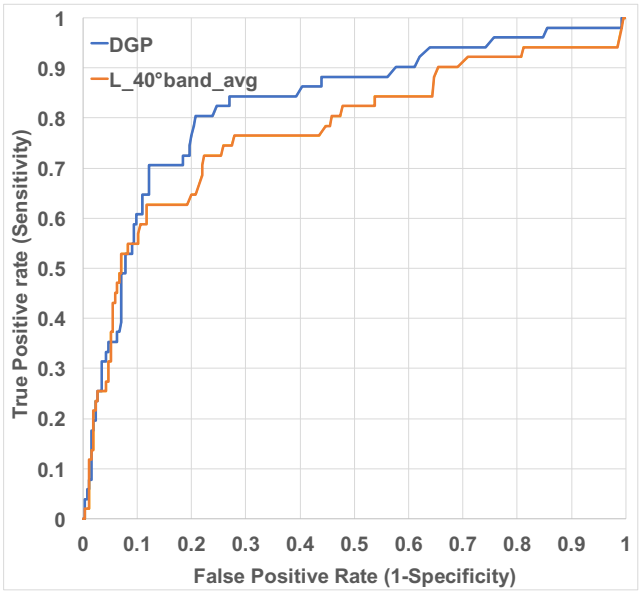
The Bonferroni-Correction [13] of significance values is in general applied when multiple statistical tests of a hypothesis are performed. If a test is applied multiple times, then the probability that one of the tests is randomly positive increases. The Bonferroni-correction compensates this by adjusting the significance level α by the amount of tests applied. In our study we investigate twelve metrics at the same time, therefore the significance levels have to be adjusted according to table 2. The adjusted levels from table 2 are used for all statistical tests in this study.

Table 2 – Bonferroni-adjusted significance levels.

Significance Level	α_1 (one test)	$\alpha = \frac{\alpha_1}{12}$
*	0.05	0.0042
**	0.01	0.00083
***	0.001	0.000083

3.3.3 ROC curve analysis

The ROC-curve (Receiver Operating Characteristic curve) is a tool to evaluate (diagnostic) tests. Usually it is applied in medical studies to evaluate and compare different testing methods, but was introduced by [11] to glare evaluations. It is applied to binary, dependent variables. Therefore, we converted for our study the subjective, ordinal data on the 4-point-likertscale to binary data (not disturbed ↔ disturbed by glare). Several indicators can be derived from this curve, but for our study we derived two values:



1. The cut-off-value of each metric (training data). The cut-off-point of a metric is defined as the borderline of the metric, dividing "not-disturbed" from "disturbed". For our study, the optimum cut-off-point is derived by minimizing the distance to the upper left corner (Sensitivity=1, Specificity=0).

2. The area under the ROC-curve (AUC): The larger the area, the better is the prediction of the metric. (applied to the non-dev dataset)

The ROC curve is defined by the sensitivity (or also called true positive rate) on the y-axis and 1-Specificity (also called false positive rate) on the x-axis. Each point on that curve is calculated by another cut-off-point.

Figure 3 – An example of ROC curves for two of the metrics.

3.3.4 Accuracy analysis

The accuracy of a metric is defined as the fraction of true predictions of the model out of the total amount of cases.

$$accuracy = \frac{true\ predictions}{total\ cases} \tag{1}$$

For binary data, this value must be larger than 50%, otherwise a purely random model would have a similar prediction rate.

In this study, the cut-off-values for each metric was first derived from the training dataset, including the development data from DGP in order to have a very broad training dataset. The accuracy calculation does not include any training data and is applied to the testing dataset.

3.3.5 Logistic regression

The logistic regression is a regression method for binary dependent variables. The outcome of the regression are the coefficients a and b for the equation (2). For this study, we only evaluate the p-values of the logistic regression.

$$P = \frac{e^{a+bx}}{1+e^{a+bx}} \tag{2}$$

3.3.6 Analysis of the probability of being disturbed by glare

This analysis was introduced to glare analysis in [2]. For this evaluation, the analysed data are first sorted by the metric values and then binned into ten bins. The number of bins is set to ten based on typical binning numbers when performing a Hosmer-Lemeshow-test [20] on logistic regressions. However, the choice of the number of bins is arbitrary and influences the results, which is the main reason the Hosmer-Lemeshow test is under discussion [21]. On the other hand, the comparison between different metrics (in our case based on r^2 values) using this method provides an informative assessment of the performance of the metric, since this test is intuitively evaluating the probability of users being disturbed by glare and relating it to the metric.

For each bin, the probability of being disturbed by glare is calculated from the subjective response as well as the average metric value. For the resulting binned data (probability of being disturbed by glare ↔ average metric) a linear regression is applied and r^2 and p are calculated.

4 Results

The evaluation of the data aims to determine how different discomfort glare metrics perform when they are applied to datasets other than those they emerged from. Unfortunately, as mentioned in chapter 3.3, the performance analysis of a glare metric is not uniquely defined. Therefore, several methods had to be applied here to answer the main two questions of the study:

1. Ability for a metric to describe the full glare scale: for this, Spearman correlations (and their p-values) between the metrics and the glare scale are calculated.

2. Accuracy or predictive ability of the metric: for this, logistic regressions, accuracy analysis and probability correlations are applied to the data-sets.

4.1 Analysis of glare metrics vs. perceived glare magnitude

In table 3 and figure 4 the results of the Spearman analysis are shown. Except for the average luminance of the window (not significant for the Ecco-Build data set), all other metrics reach the ** (“very significant”) or *** (“extremely significant”) significance levels. The levels of correlation differ between the studies (highest values for the study “Quanta”, lowest values for the study “Ecco-Build”) but also between the metrics. The level of the correlation values gives the first indication of the performance of the metrics when applied to conditions that were not part of the metric’s development (so-called non-dev-data). More specifically, the non-dev dataset correlations show that the highest correlations were found for DGP (ρ=0.485) and CGI (ρ=0.477) whereas E_v, $L_{40°band\_avg}$, L_{avg}, L_{avg_win} and $pgsv_{sat}$ show slightly lower correlations (ρ=0.41-0.42). The other metrics are in-between.

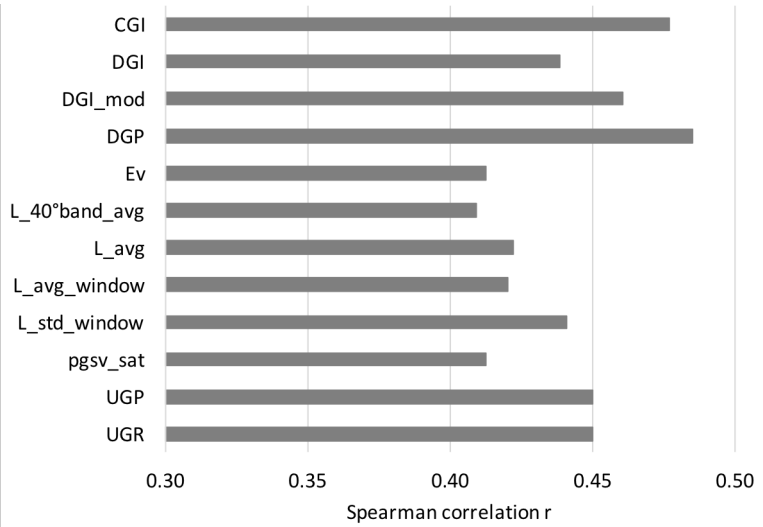


Figure 4 – Spearman correlation between metrics and the subjective glare response

Looking separately on each study-dataset one can see, that the ranking between the different metrics change (for the Quanta-dataset DGP, Ev, L_{avg},and pgsv showed the highest correlations, whereas for the Gaze-data-set it is DGI, UGP and UGR).

Table 3 – Spearman correlation evaluation for different data-sets (for Ecco-Build, Quanta and Gaze all data are included)

	Spearman r-value				Level of significance			
	Ecco-Build	Quanta	Gaze	non-dev-data	Ecco-Build	Quanta	Gaze	non-dev-data
CGI	0.29	0.53	0.41	0.48	***	***	***	***
DGI	0.27	0.46	0.43	0.44	***	***	***	***
DGI_mod	0.27	0.50	0.41	0.46	***	***	***	***
DGP	0.33	0.55	0.38	0.48	***	***	***	***
Ev	0.29	0.55	0.38	0.41	***	***	**	***
$L_{40°band\_avg}$	0.26	0.53	0.35	0.41	***	***	**	***
L_avg	0.26	0.55	0.39	0.42	***	***	***	***
L_avg_window	0.08	0.54	0.39	0.42	-	***	***	***
L_std_window	0.21	0.50	0.40	0.44	***	***	***	***
pgsv_sat	0.29	0.55	0.38	0.41	***	***	**	***
UGP	0.25	0.49	0.43	0.45	***	***	***	***
UGR	0.25	0.49	0.43	0.45	***	***	***	***

NOTE The Ecco-Build-dataset contain development data for the DGP, therefore the correlation value for the DGP cannot be compared directly with the values from the other metrics.

As conclusion from the Spearman evaluation it can be said, that the relative differences between the correlations of the different metrics are within 15% and the ranking is changing when evaluating different data-sets. Therefore, for most of the applied metrics the Spearman correlation shows no clear preference of one of the metrics. The only exception is L_{avg_window} which is failing one of the significance tests for one of the data-sets.

4.2 Analysis of glare metrics vs. probability of being disturbed by glare

For this analysis, the subjective response data are converted into binary data (not disturbed by glare ↔ disturbed by glare).

Furthermore, the entire data is split into a “training dataset” (for deriving the regression coefficients of the logistic regression and the cut-off-points for the accuracy calculation) and into a “testing dataset” (details described in chapter 3.2).

Applying the logistic regression (see table 4), the significance test is failed for three metrics for at least one of the data-sets (the average luminance of the 40° band $L_{40°band\_avg}$, the average luminance of the window L_{avg_window} and the standard deviation of the window luminance L_{std_window}). This means, that a logistic regression model for these metrics is not a good choice and therefore it can be assumed that these metrics cannot predict reliably the percentage of persons disturbed for any dataset.

For the accuracy evaluation, the cut-off-point for each metric was derived from the “training dataset”, consisting data of all studies. These cut-off-points are used to calculate the accuracy of the metrics for the “testing dataset”. The accuracy is defined as fraction of correct predictions by the metric for the binary data. Two metrics are failing a minimum prediction threshold of 50% for one of the datasets (L_{avg_window} and L_{std_window}, see table 4 and figure 5). The analysis of area under the ROC-curve (AUC) show, that there are small differences between the metrics. The DGP, CGI, UGP and DGI_mod have the highest values, indicating a slightly better performance.

It can be concluded from this analysis that besides L_{avg_window} and L_{std_window} the other metrics are behaving very similar. $L_{40°band\_avg}$ is failing one test. The accuracy levels vary more between the studies than between the metrics. The Ecco-Build-study consists of many situations with blinds diffusing the light. These situations were rated very differently by the people and therefore the noise of the data is larger than for other data-sets, where people were more consistently rating. Therefore, the accuracy is also lower for this study. For the combined dataset (“testing-non-dev-dataset”), the accuracy of the metrics-predictions is in the range of 75-83%.

Table 4 – Significance test for the logistic regression and accuracy analysis

	Logistic regression - level of significance			Accuracy Disturbed				AUC
	Ecco-Build	Quanta	Gaze	Ecco-Build	Quanta	Gaze	Non-Dev	Non-Dev
CGI	***	***	***	0.56	0.76	0.87	0.79	0.83
DGI	**	***	**	0.58	0.70	0.86	0.74	0.81
DGI_mod	***	***	**	0.56	0.78	0.86	0.81	0.82
DGP	***	***	**	0.59	0.74	0.84	0.79	0.83
Ev	***	***	**	0.60	0.74	0.81	0.78	0.79
L_40°band_avg	***	***	-	0.58	0.76	0.89	0.83	0.78
L_avg	***	***	***	0.62	0.75	0.84	0.81	0.79
L_avg_window	-	***	***	0.49	0.81	0.86	0.84	0.79
L_std_window	***	-	***	0.49	0.78	0.76	0.78	0.80
pgsv_sat	**	***	*	0.60	0.74	0.81	0.78	0.79
UGP	*	***	**	0.58	0.73	0.86	0.75	0.82
UGR	*	***	**	0.58	0.73	0.86	0.75	0.76

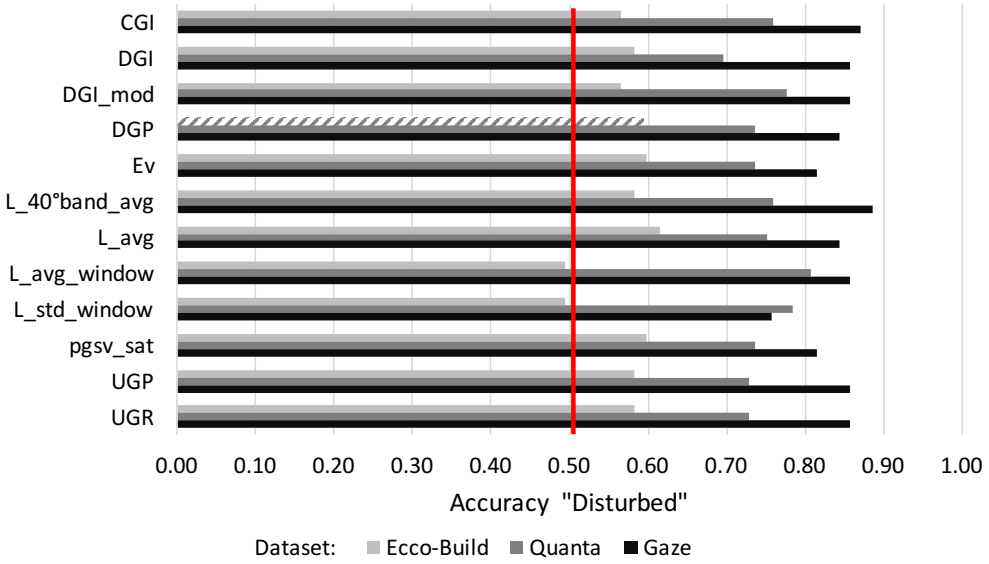


Figure 5 – Accuracy of the metrics for the three studies.

The final statistical analysis is comparing the ability of a metric to predict directly the probability of persons disturbed by glare. This evaluation does not include data from the DGP-development.

Table 5 – Squared Pearson correlation between metric and probability of being disturbed

Metric	r^2	Level of significance	Metric	r^2	Level of significance
CGI	0.68	*	L_avg	0.77	*
DGI	0.64	-	L_avg_window	0.66	*
DGI_mod	0.73	*	L_std_window	0.65	-
DGP	0.89	***	pgsv_sat	0.43	-
Ev	0.84	**	UGP	0.73	*
L_40°band_avg	0.80	**	UGR	0.73	*

The data is sorted by the metric value and binned into ten bins (see chapter 3.3.6). For each of the bins, the average metric value and for the subjective response the probability of being disturbed is calculated. Then, a linear regression is performed. The r^2-value itself cannot be interpreted absolutely, since the number of bins is influencing this value strongly. But the difference between the r^2-values of the different metrics and the shape of the regression graphs for the different metrics gives an indication about how well a certain metric can be used to predict the probability of being disturbed by glare.

The analysis show (see table 5 and figure 6), that DGP, Ev and $L_{40°band\_avg}$, have higher levels for r^2 and higher levels of significance than the other metrics. L_avg, UGP, DGI_mod, CGI and UGR predict still reasonably the probability of persons disturbed, whereas DGI, pgsv_sat, L_{avg_window} and L_{std_window}, are metrics failing the significance test and showing low correlation with the probability of being disturbed by glare.

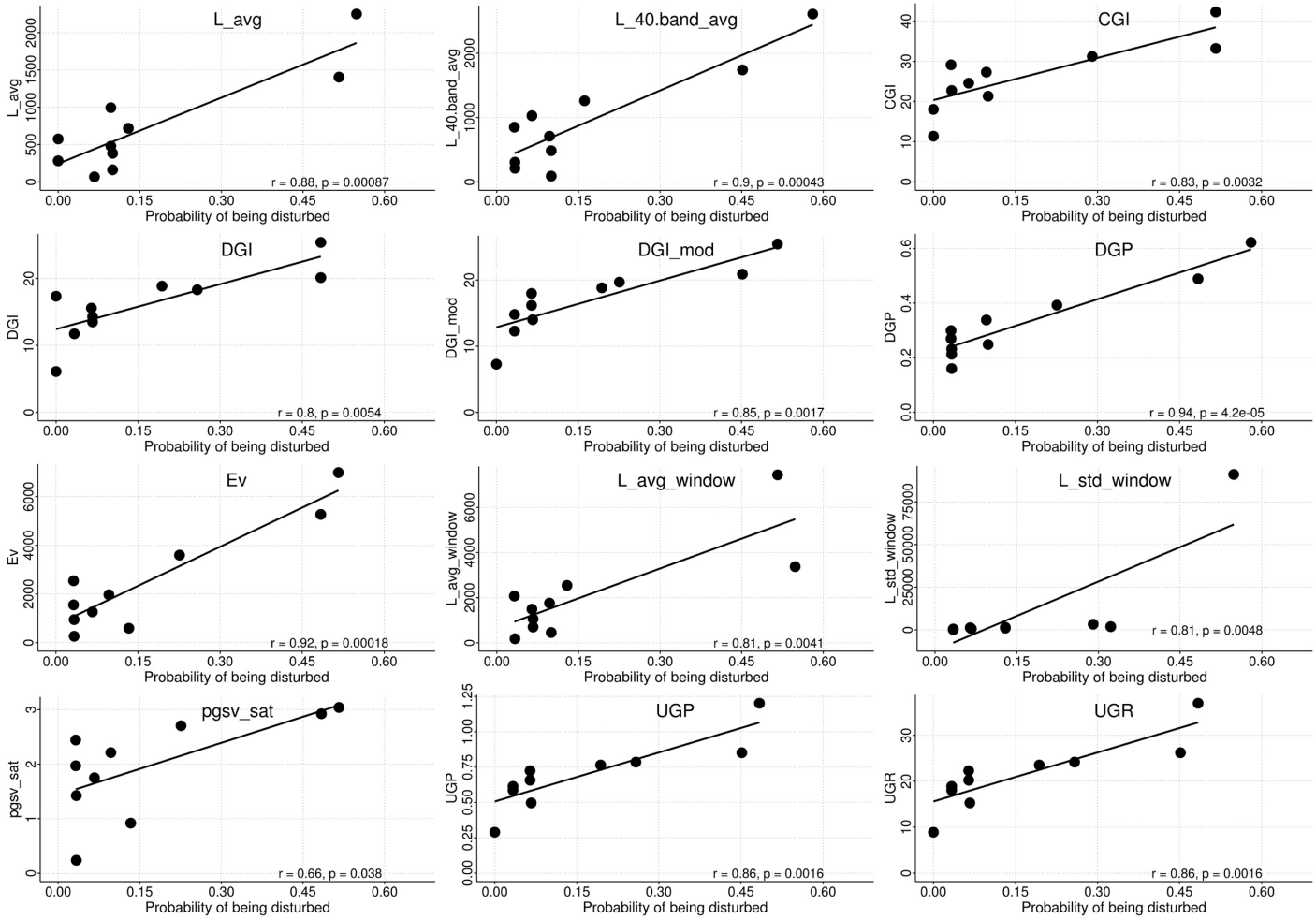


Figure 6 – Linear regressions between average metric value and probability of being disturbed by glare. The data are binned into ten binning groups in order to calculate the probability of being disturbed by glare

5 Conclusions and outlook

In this paper we evaluated twelve glare metrics regarding their ability to predict the glare perception of subjects of three laboratory assessments.

The analysis shows that five of the twelve investigated metrics fail at least one statistical test for at least one of the datasets. The other seven metrics CGI, DGI_mod, DGP, Ev, Lavg, UGP and UGR pass all statistical tests. For the Spearman and accuracy evaluations, the difference in the performance of these seven metrics is small and therefore no clear superiority or inferiority of one of the metrics can be concluded from this. The accuracy of the metrics' predictions is in the range of 75-83%. Therefore, this study cannot confirm in general a poor performance of existing glare metrics when applied to different datasets than those they were developed with.

Amongst the seven non-failing metrics, DGP, CGI, DGI_mod and UGP have the largest AUC and therefore show a slightly better ability to predict the probability of being disturbed by glare than the other metrics. For the linear regression between probability being disturbed by glare and value of a respective metric, DGP and EV perform better than the other metrics.

From this study it can be concluded, that not all the metrics can predict the glare perception with the same reliability when applied to different daylighting conditions and that this needs further evaluation. The usage of one of the metrics failing one or more statistical tests should be done only with caution. However, given the similarity of the experimental setups and the limited geographical variation in this study (two locations within Europe, three studies in total) general conclusions regarding robustness of the metrics or recommendations for their usage under specific conditions would be premature. Also, proposing robust cut-off-values for the metrics would need to be based on broader data-sets.

A more comprehensive comparison based on additional studies conducted on different continents is under preparation and might answer these open questions.

References

1. Hopkinson, R.G. Glare from daylighting in buildings. Applied Ergonomics 1972
2. Wienold, J,. Christoffersen, J. Evaluation methods and development of a new glare prediction model for daylight environments with the use of CCD cameras. Energy and Buildings,38(7): 743-757, 2006.
3. Sarey Khanie, M., Stoll, J., Einhäuser, W., Wienold, J., Andersen, M., "Gaze and discomfort glare, Part 1: Development of a gaze-driven photometry", Lighting Research and Technology, 2016.
4. Fisekis, K., and Davies, M. Prediction of discomfort glare from windows. Lighting Research and Technology 35, 360–371, 2003.
5. K. Van Den Wymelenberg and M. Inanici, “A Critical Investigation of Common Lighting Design Metrics for Predicting Human Visual Comfort in Offices with Daylight,” LEUKOS, vol. 10, no. 3, pp. 145–164, 2014.
6. K. Van Den Wymelenberg and M. Inanici, Evaluating a New Suite of Luminance-Based Design Metrics for Predicting Human Visual Comfort in Offices with Daylight, LEUKOS, 2015.
7. Hirning, M.B., Isoardi, G.L., and Cowling, I. Discomfort glare in open plan green buildings. Energy Build. 70, 427–440, 2014.
8. Boyce P.R. Human factors in lighting - third edition. ISBN 9781439874882. CRC Press, 2014
9. Galasiu A. D., Veitch J. A., Occupant preferences and satisfaction with the luminous environment and control systems in daylit offices: a literature review, Energy and Buildings, Volume 38, Issue 7, Pages 728-742, 2006
10. Wittwer, V., Wagner, A., Moosmann, C., Wienold, J., Ermittlung relevanter Einflussgrößen auf die subjektive Bewertung von Tageslicht zur Bewertung des visuellen Komforts in Büroräumen. Final report. URN: urn:nbn:de:swb:90-349684, Karlsruhe, 2012.
11. Rodriguez R. G. , Garretón, J.A.Y., Pattini A. E., An epidemiological approach to daylight discomfort glare, Building and Environment, Volume 113, Pages 39-48, 2017
12. Siegel S., Nonparametric Statistics for the Behavioral Sciences. 1st edition. New-York: McGraw-Hill, 1956
13. Bonferroni, C. E., Teoria statistica delle classi e calcolo delle probabilità, Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze, 1936
14. Einhorn, H.D., A New Method for the Assessment of Discomfort Glare. Lighting Research and Technology, 1, 235-247,1969.
15. Einhorn, H.D., Discomfort Glare: A Formula to Bridge Differences. Lighting Research and Technology, 11, 90-94, 1979.
16. CIE, Discomfort Glare in the Interior Lighting. T.c.T.-. Commission Internationale de l’Éclairage (CIE), Division 4,1992
17. Tokura, M., Iwata, T. and Shukuya, M., Experimental Study on Discomfort Glare Caused by Windows, Part 3. Development of a Method for Evaluating Discomfort Glare from a Large Light Source, Journal of Architecture, Planning and Environmental Engineering., 489, 1996
18. Iwata, T., Nagayoshi, K., Osterhaus, W., Discomfort glare index for automated blind control, Proceedings of the ISIS Solar World Congress 2011, Kassel, Germany, 2011
19. Wienold, J., Evalglare 2.0 – new features, faster and more robust HDR-image evaluation, 15th International Radiance Workshop, Padova, Italy, 2016
20. Hosmer, David W.; Lemeshow, Stanley (2013). Applied Logistic Regression. New York: Wiley.
21. Bertolini G, D'Amico R, Nardi D, Tinazzi A, Apolone G, et al., One model, several results: the paradox of the hosmer–lemeshow goodness-of-fit test for the logistic regression model. J Epidemiol Biostat 5: 251–253, 2000